CSMC828G | Principles of Data Mining | Spring 2002

# Principles of Data Mining
## Course Logistics

**Web page:** http://www.cs.umd.edu/class/spring2002/cmsc828g/

**Time and Place:** Tu,Th 11:00-12:15 AVW 3258

**Instructor:** Lise Getoor, AVW Bldg, Rm. 3205, getoor@cs.umd.edu
office hours: Tu 3:30-5:30PM and by appointment

**Teaching Assistant:** Eiman Einahrawy, AVW Bldg, Rm. 3228, eiman@cs.umd.edu
office hours: Tu,Th 1:30-2:30PM and by appointment

**Course Description:** This course covers the fundamentals of data mining from both a statistical and database perspective. The course has three major sections. The first section of the course will cover the statistical and machine learning foundations for data mining. In the second section of the course, we will cover the fundamental data mining concepts and algorithms for tasks such as OLAP, association rules, clustering, etc. The final section of the course will focus on research areas such as text mining, collaborative filtering, link analysis and mining in biological domains (as time permits).

**Prerequisites:** CMSC 421, Introduction to Artificial Intelligence and CMSC 424, Database Design, or equivalent courses.

**Useful background:** Probability theory, statistics.

**Other useful courses:** CMSC 726, Machine Learning, CMSC 838L, Information Retrieval Systems, and CMSC 724, Database Management Systems.

**Workload:** There will be an in-class midterm and final. There will be three homeworks which will include a written portion and a computer portion (typically using existing matlab infrastructure). A major component of the workload will be a class project.

**Grading:** Midterm (20%), Final (30%), Project (35%), HW (15%).

**Mailing list:** cmsc828g@cs.umd.edu. Please subscribe by sending mail to majordomo@cs.umd.edu with 'subscribe cmsc828g' in the body of the message.

**Text:**
Principles of Data Mining by David J. Hand, Heikki Mannila and Padhraic Smyth, August 2001.

**Other supplementary texts:**
Predictive Data Mining : A Practical Guide by Sholom M. Weiss, Nitin Indurkhya. Morgan Kaufmann Publishers, 1997.

Data Mining: Concepts and Techniques by Jiawei Han, Micheline Kamber. Morgan Kaufmann Publishers, 2000.

Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations by Ian H. Witten, Eibe Frank. Morgan Kaufmann Publishers, 1999.

Advances in Knowledge Discovery and Data Mining by Usama M. Fayyad , Gregory Piatetsky-Shapiro, Padhraic Smyth, Ramasamy Uthurusamy. MIT Press, 1996.

The Elements of Statistical Learning: Data Mining, Inference, and Prediction by Trevor Hastie, Rob Tibshirani and Jerome Friedman. Springer Verlag Publishers, 2001.